IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 June 2024, accepted 8 July 2024, date of publication 12 July 2024, date of current version 22 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3427146

RESEARCH ARTICLE

A Deep Learning-Based Fault Diagnosis Method for Flexible Converter Valve Equipment

JIANBAO GUO[1], HANG LIU[1], LEI FENG[1], LIFENG ZU[2], TAIHU MA[2], AND XIAOLE MU[2]
[1]EHV Maintenance & Test Center of China Southern Power Grid, Guangzhou 510663, China
[2]XJ Electric Flexible Transmission Company, Xuchang 461000, China

Corresponding author: Taihu Ma (mtaihu@163.com)

This work was supported by the Innovation Project of China Southern Power Grid Company Ltd., under Grant CGYKJXM20220059.

ABSTRACT Long-term failures in flexible converter valve equipment pose significant risks, potentially compromising operational efficiency or leading to complete malfunction. Accurately identifying equipment faults is essential to improve overall reliability and minimize downtime. This study introduces an innovative fault diagnosis method utilizing an attention mechanism. The method integrates a lightweight model incorporating one-dimension depthwise convolutional layers for spatial feature extraction and bidirectional long short-term memory for capturing temporal dynamics. A pioneering time-channel joint attention module enhances the extraction of fault-related data from time series and channel maps. Experimental results underscore the method's efficacy in fault diagnosis under varying Gaussian noise conditions. Notably, the approach demonstrates remarkable consistency in accuracy across various experimental setups, underscoring its robust performance and potential applicability in real-world scenarios where reliability is critical. In addition, the proposed method has a moderate number of parameters and training time, indicating that the model can be embedded in front-end equipment.

INDEX TERMS Bidirectional long short-term memory, channel attention module, deep learning, depth-wise convolution, fault diagnosis, flexible converter valve equipment, high voltage direct current system, lightweight, power system, time attention module.

NOMENCLATURE

DC	Direct current.
SVM	Support vector machine.
PCA	Principal component anlysis.
RF	Random forest.
CNN	Convolutional neural network.
1DCNN	One-dimensional CNN.
GRU	Gated recurrent unit.
LSTM	Long short-term memory.
BiLSTM	Bidirectional LSTM.
$\text{Linear}_i(\cdot)$	Fully connected layer.
δ	ReLU function.
σ	Sigmoid function.
$f^{1\times k}$	Filter with size $1 \times k$.
$\overline{X}$	Average value of X.
X_{gmp}	Global maximum pooling.
X_{gap}	Global average pooling.
SCE	Secure crypto engine.
TP	True positive.
TN	True negative.
FP	False positive.
FN	False negative.
BPNN	Back propagation neural network.
DSC	Depth-wise convolution.
SE	Squeeze-and-excitation.
CBAM	Convolutional block attention module.
JAM	Joint attention module.
TCJAM	Time-channel joint attention module.
SM	Sub-module.
i_t	Input gate.
f_t	Forget gate.
o_t	Output gate.
$\widetilde{C}_t$	Candidate cell state.
C_t	Cell state.

The associate editor coordinating the review of this manuscript and approving it for publication was Gerard-Andre Capolino.

h_t	Hidden state.
$\mathbb{R}^{N\times d}$	Input and output samples with dimensions of N and d.
$\overrightarrow{\text{LSTM}}(X)$	Forward LSTM.
$\overleftarrow{\text{LSTM}}(X)$	Backward LSTM

I. INTRODUCTION

The flexible converter valve equipment is crucial in High Voltage Direct Current (HVDC) systems, facilitating the conversion between HVDC transmission and alternating current transmission [1]. However, the flexible converter valve equipment is prone to failure after long-term operation due to its complex structure and extreme operating conditions, which affects the normal operation of HVDC [2]. Therefore, developing an effective fault diagnosis method is of great significance for ensuring the safe and stable operation of the power grid [3].

Due to the limited research on fault diagnosis for flexible converter valve equipment, we conducted research on HVDC fault diagnosis as a reference. Current fault diagnosis methods on HVDC fault diagnosis are primarily based on models and machine learning [4]. Model-based methods mainly rely on empirical judgment or simple rules for fault diagnosis, examples of which include fault observer [5], expert systems [6], [7], fault tree analysis [8], and model prediction methods [9]. Although these methods can infer the causes of faults and locate faulty components from the perspective of fault occurrence mechanisms, they have limitations such as subjectivity, low diagnostic efficiency, and susceptibility to human factors. In addition, model-based methods have limited capabilities in handling complex and diverse fault patterns and large amounts of data, thus not satisfying practical application requirements [10]. The complex and diverse nature of the flexible converter valve equipment limits the performance of model-based methods in feature extraction and rule design, and their significant manual involvement requirements render them unsuitable for this context.

Accordingly, several fault diagnosis methods for flexible converter valve equipment have been proposed based on machine learning algorithms [11], [12], [13], [14]. By learning the operational data generated during the operation of flexible converter valve equipment (e.g., power signals), machine learning algorithms can obtain the mapping relationship between power signals and different fault states without relying on expert knowledge. For example, Ghashghaei et al. [11] adopted Support Vector Machine (SVM) and K-Nearest Neighbours (KNN) algorithms to detect DC transmission line faults and serve as a redundant module for unsure fault declaration from the startup unit. Similarly, Ye et al. [12] and Zhou et al. [15] adopted different SVM kernel functions to quickly and accurately diagnose faults in simulated circuits. Under the ensemble learning framework, Movahed et al. [13] proposed a method based on Principal Component Analysis (PCA) and Random Forest (RF) to reduce dataset dimension and account for data imbalance, thus reducing misleading alarms in fault diagnosis. Under the artificial neural network framework, Liu et al. [14] proposed a Backpropagation Neural Network (BPNN) method for flexible DC transmission line fault diagnosis, which achieved high-precision fault localization by analyzing operation mode and control strategy. The traditional machine learning methods above use historical data and algorithm rules to iteratively derive appropriate model parameters in an automated and intelligent manner, i.e., through ''trial and error'' with multiple training episodes [16]. However, traditional machine learning methods perform poorly in the case of large amounts of data and heterogeneous and complex signal diagnostics.

Recently, deep learning-based fault diagnosis methods have become a research hotspot due to their ability to extract fault features and achieve high diagnostic accuracy [17]. For instance, Liu et al. [18] presented a wavelet neural network based commutation failure diagnosis method, which extract fault feature by wavelet transform. Compared with the BPNN method, this method achieves higher accuracy and diagnostic speed under multiple fault types. Similarly, Wang et al. [19] proposed a flexible DC distribution system fault diagnosis method based on wavelet transform and CNN. Firstly, this method decomposes the fault voltage signal through wavelet transform and reconstructs it into a two-dimensional time-frequency image. Subsequently, the method inputs image data into the CNN model for training and testing. Although CNN has high accuracy and robustness, the data processing process is time-consuming and requires high system computing resources. Considering the electrical signal sample is a kind of temporal signals, Han et al. [20] adopted improved Long Short-term Memory (LSTM) to process the wavelet entropy fault information in the time dimension and obtained the adaptive classification results by SVM. This method can significantly reduce the number of voltage signal samples required for diagnosis. To further improve the accuracy of fault diagnosis, Zhou et al. [21] integrated multi-layer perceptron, LSTM, and CNN models to construct a weighted fault diagnosis model, accurately diagnosing the Converter Valves device. However, this method has a large number of parameters and high complexity, and has limitations in resource limited scenarios. The LSTM method can effectively capture and remember dependency relationships in long sequences, but it is computationally complex and has high training and inference costs. After adding bidirectional information flow, the BiLSTM method considers both past and future contextual information, which helps to predict the current state or fault more accurately. However, compared to unidirectional LSTM, BiLSTM requires more computing resources and time for training and inference.

To reduce model complexity, Xia et al. [22] used a GRU model with a simpler structure and faster training speed for fault diagnosis of wind turbines. Although GRU has fewer gating units and fewer parameters, it cannot effectively handle certain complex time series patterns. To further reduce model

complexity, Liu et al. [23] proposed a shallow deep belief network model based on parameter optimization, which reduces the time for fault signal processing while ensuring diagnostic accuracy. In addition, Hou et al. [24] proposed a fault diagnosis method based on depthwise separable convolution, with a parameter size of only 1.912 KB and inference time of only 75% of that of ordinary one-dimensional neural networks. It should be noted that although depthwise separable convolution can significantly reduce model parameters, it may to some extent lose the perception ability of global information.

Considering the significant impact of faults in flexible converter valve equipment on voltage signals, it is beneficial to refer to the above methods for accurately identifying abnormal changes in voltage signals. However, the prerequisite of expanding the above methods to the flexible converter valve equipment lies in the analysis and summary of the challenges in its fault diagnosis, which are as follows:

- The equipment has a complex structure and variable working environment, and the resultant non-linear characteristics in sensor data increase the fault diagnosis difficulty.
- Model-based methods are based on manual feature extraction and rule design, involving the extensive utilization of the complex characteristics of flexible converter valve equipment.
- Since earlier fault diagnosis is more conducive to repair work, simplifying the fault diagnosis model and deploying it on the device side can improve the timeliness of fault diagnosis, which has high requirements for model design.

This study proposes a fault diagnosis method for flexible converter valve equipment based on a one-dimensional convolutional neural network (1DCNN) and an attention mechanism in deep learning, aiming to address the above research challenges. The specific contributions are as follows:

- A fault diagnosis model based on 1DCNN and an attention mechanism is proposed, which accurately diagnoses faults in flexible converter valve equipment by automatically learning feature representation and rule discrimination capabilities.
- The number of samples for small-sample-size categories is expanded through overlapping sampling, thereby avoiding ignoring the small-sample-size categories during training.
- The effectiveness and feasibility of the proposed method are validated through experiments and comparison with traditional methods, demonstrating the advantages of the proposed method in the fault diagnosis of flexible converter valve equipment.

This work aims to improve the accuracy and efficiency of fault diagnosis for flexible converter valve equipment using deep learning and attention mechanisms, providing technical support for safe and stable power grid operation.

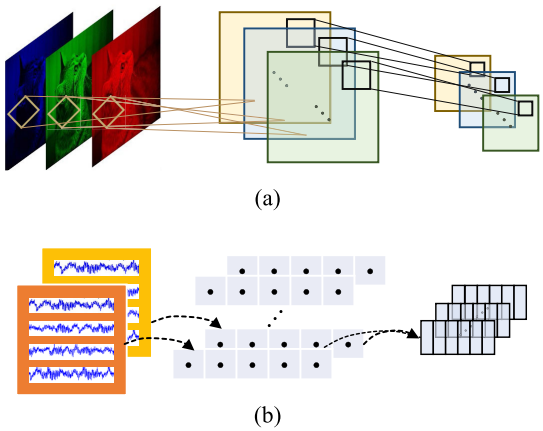

FIGURE 1. Different input features targeted by (a) Ordinary convolution and (b) One-dimensional convolution.

II. BASIC THEORY

A. ONE-DIMENSIONAL CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Networks (CNNs) are artificial neural networks specifically designed to process video and image data. CNNs extract features from input images and learn to classify output images based on the learned features [25]. As shown in Figure 1 (a), ordinary convolution, often referred to in the context of two-dimensional convolution, is widely used in image processing. This process involves sliding a kernel (or filter) over the image (or input data) in two dimensions (height and width). At each position, the kernel values are multiplied with the values of the input pixels that they cover, and the results are summed to produce a single pixel in the output feature map. This operation helps to detect features such as edges, textures, or gradients in images. In order to handle numerical data, the 1DCNN has been developed and applied in signal processing and sequence data analysis. As shown in Figure 1 (b), 1DCNN performs convolution operations on input sequences to extract features and applies them to tasks, e.g., classification and regression of sequence data. In this type, convolutional operations are performed along a single spatial dimension (the time axis, in the case of time-series data). 1DCNN has several important characteristics:

First, 1DCNN can capture local correlations within the input sequence. By defining convolutional kernels of different sizes, 1DCNN can perform sliding window operations on the input sequence at different scales and extract local subsequences of different lengths, which contributes to effectively capturing the local patterns and features within the input sequence. Secondly, 1DCNN has a parameter sharing mechanism. Each convolutional kernel performs convolution operations with the entire input sequence in the convolutional layers, thereby generating new feature maps. Thus, the number of parameters the model needs to learn is not affected by the length of the input sequence but remains constant. This parameter-sharing mechanism significantly reduces model complexity and enhances training efficiency.

Furthermore, 1DCNN can increase the depth and complexity of the model by stacking multiple convolutional layers and pooling layers. Through hierarchical feature extraction and abstraction, 1DCNN gradually learns more advanced and abstract feature representations, enhancing the model's expressive power in the case of complex sequence data.

B. BIDIRECTIONAL LONG SHORT-TERM MEMORY NETWORK

The bidirectional long short-term memory network (BiLSTM) is a sequence model that uses three gate units to address the gradient vanishing problem in traditional recurrent neural networks [26]. When a given time series is input into the LSTM layer, the LSTM unit computes the t-th data in (1) [27]:

$$\begin{cases} i_t = \sigma(\text{Linear}_i([h_{t-1}, x_{t-1}])) \\ f_t = \sigma(\text{Linear}_f([h_{t-1}, x_{t-1}])) \\ o_t = \sigma(\text{Linear}_o([h_{t-1}, x_{t-1}])) \\ \widetilde{C}_t = \tanh(\text{Linear}_c([h_{t-1}, x_{t-1}])) \\ C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \\ h_t = o_t \odot \tanh(C_t) \end{cases} \tag{1}$$

where $Linear_i(\cdot)$ represents fully connected layers, and $k \in \{i, f, o, c\}$ represents the input gate, forget gate, output gate, and cell state, respectively. $H = \{h_1, h_2, \cdots, h_N\}$ represents all the time outputs of the encoder. The input gate i_t regulates the amount of new information that should be stored in the cell state. This gate inputs the concatenation results of the previous hidden state (h_{t-1}) and the current input (x_{t-1}) into $Linear_i(\cdot)$, and then applies σ to calculate. The forget gate (f_t) determines how much information is retained from the previous cell state. This gate is also calculated through $Linear_i(\cdot)$ and σ. The output gate (o_t) controls how much information in the current cell state will be transmitted to the next layer. It is calculated through $Linear_i(\cdot)$ and σ. The candidate cell state ($\widetilde{C}_t$) using the tanh activation function to process the output of $Linear_i(\cdot)$, it is also calculated based on the hidden state of the previous time and the current input. Cell state (C_t) updated by the product of the weights of f_t and i_t, as well as $\widetilde{C}_t$. Hidden state (h_t) calculated based on o_t and the updated C_t.

As expressed in (2) [28], BiLSTM comprises two LSTM networks with opposite directions, where N and d represent the length of the input samples and the dimension of the output features, respectively.

$$H = \text{BiLSTM}(X) \in \mathbb{R}^{N \times d} = [\overrightarrow{\text{LSTM}}(X); \overleftarrow{\text{LSTM}}(X)]' \tag{2}$$

The forward LSTM and the backward LSTM process the input sequence data in chronological order and reverse order, respectively, as shown in Figure 2. This setup enables BiLSTM to simultaneously capture both preceding and succeeding temporal information around the current time step. By concatenating or merging the forward and backward

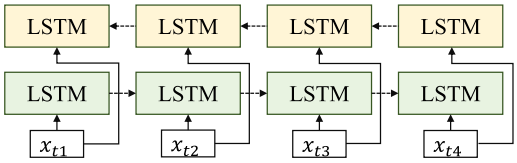

FIGURE 2. Bidirectional long short-term memory network simultaneously processing data in forward and reverse orders.

LSTM outputs, a comprehensive representation can be obtained that considers the entire input sequence information. This bidirectional model design allows BiLSTM to better capture and reveal the various patterns and regularities underlying the temporal data.

C. ATTENTION MECHANISM

The attention mechanism enhances the model's focus on important parts of the input sequence [29]. In sequence data processing tasks, it helps the model automatically learn the contribution of different positions or features to the task and assigns weights to them according to their importance. It also allows the model to treat different parts of the input sequence more flexibly. In fault diagnosis tasks, the attention mechanism enables the model to direct more attention to fault-related signal segments or features, thereby improving the accuracy and robustness of fault diagnosis.

As shown in Figure 3, the attention module can stimulate fault-related features from the periodic regularity of time-series data or the importance of different convolutional kernel channels. The time attention module is used to weigh the information from different times in time series data, so that the model can identify abnormal trends in fault features over time. Usually, one-dimensional convolutional or recursive neural networks are used to model temporal dependencies. Based on this, the time attention module can dynamically allocate attention weights for different time steps, allowing the model to better focus on important time segments whereas ignoring irrelevant parts. The channel attention module assigns weights to the information of different channels so that the model can focus on the most discriminative feature channels. Usually, it obtains global information for each channel through global pooling operations such as global average pooling or global maximum pooling whereas learning the weight of each channel through fully connected layers or one-dimensional convolutional layers. On this basis, the module can dynamically adjust the importance of each channel, thus improving the model's effectiveness in feature extraction. The time-channel joint attention module combines the advantages of the time and channel attention modules to simultaneously capture the correlations between time and channels. It first weights the information of each time step through the time attention module and then weights the weighted feature map through the channel attention module. In this way, the time-channel joint attention module can dynamically adjust the attention weights in the time and channel dimensions, thereby better capturing key information in time series data.

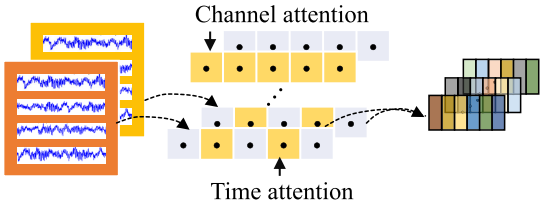

FIGURE 3. Time-channel joint attention module. The time attention module optimizes weights on time steps. The channel attention module optimizes weights between channels. Both attention modules strengthen fault-related features in different dimensions to improve the accuracy of fault diagnosis models.

In addition, the attention mechanism provides an interpretable method for understanding the decision-making process of deep learning networks. By observing the distribution of attention weights, researchers can understand the preferences of the deep learning network for fault features at different time points or frequency ranges, which is helpful for subsequent fault analysis and optimization.

III. PROPOSED METHOD

The proposed fault diagnosis method combining 1DCNN-BiLSTM with an attention module is illustrated in Figure 4. The main process is as follows:

- Dataset construction: The one-dimensional voltage signals are acquired under different fault conditions and are labeled manually. To simulate the real environment, different levels of Gaussian noise are injected into the data. Subsequently, an overlapping sampling method is adopted, which groups each class of samples into sets of 300 data points, with the overlap step determined by the number of fault samples in that class.
- Feature extraction: In the first layer (DSConv1 in Figure 4), the proposed method adopts 40 one-dimensional convolutional kernels with a size of 10×1. In the second layer (TCJAM2 in Figure 4), a time-channel joint attention module is adopted to inspire fault-related maps from the first layer. The time attention module utilizes a 1×1 convolutional kernel and 40 5×1 depth-wise separable convolutional kernels to obtain importance weights along the time dimension. Meanwhile, the channel attention module uses two 5×1 fully connected layers and two 40×1 fully connected layers to allocate weights to different channels based on their importance.

  The proposed attention module maps the input sequence X into the weighted output sequence Y, as expressed in (3).

$$Y = X + \delta(f^{1\times k}(\sigma(f^{1\times 1}(\overline{X} + \max(X))))) + \sigma(f^{1\times c}(\delta(f^{1\times c/r}(X_{gmp} + X_{gap})))) \quad (3)$$

  where δ and σ represent the ReLU and Sigmoid functions. $f^{1\times k}$ indicates the $1 \times k$ filter. $\overline{X}$ and $\max(X)$ represent the average value and maximum value of X along the channel axis. $f^{1\times c}$ and $f^{1\times c/r}$ represent fully connected operations with different hidden sizes, c and r denote the number of channels and compression ratio. X_{gmp} and X_{gap} indicate global maximum pooling and global average pooling operations of X. At the third and fourth layers (BiLSTM3 and BiLSTM4 in Figure 4), temporal features are extracted from weighted feature maps through TCJAM.
- Classifier: To simplify the model, the fully connected layer is replaced with global average pooling in the final layer (Classifier in Figure 4). Subsequently, the obtained feature maps are input into a Softmax layer for classification. The model is optimized based on the specified loss function until the iteration reaches the set number or satisfies the early termination condition.

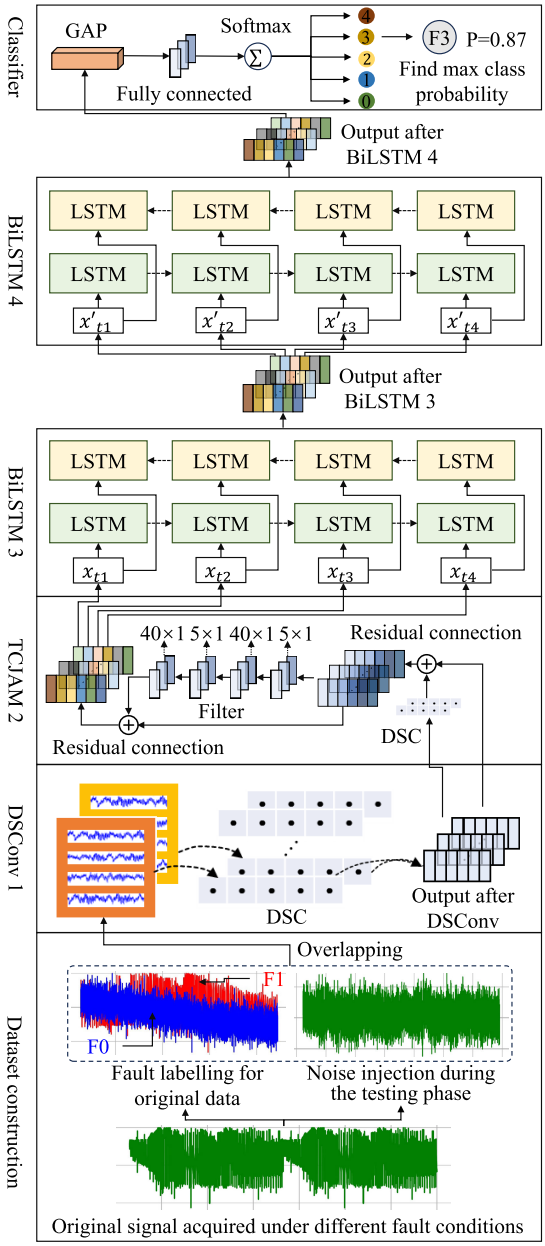

FIGURE 4. Flow chart of the fault diagnosis method combining 1DCNN-BiLSTM with attention module.

FIGURE 5. Prototype of fiber optic networking dynamic mode equipment, where 54 SCE boards are adopted as sub-modules for networking. By using few physical sub-modules can achieve full access testing of large-scale node valve control, thereby reducing the cost of power testing for the key equipment in the power system, e.g., driver boards, high-voltage power supplies.

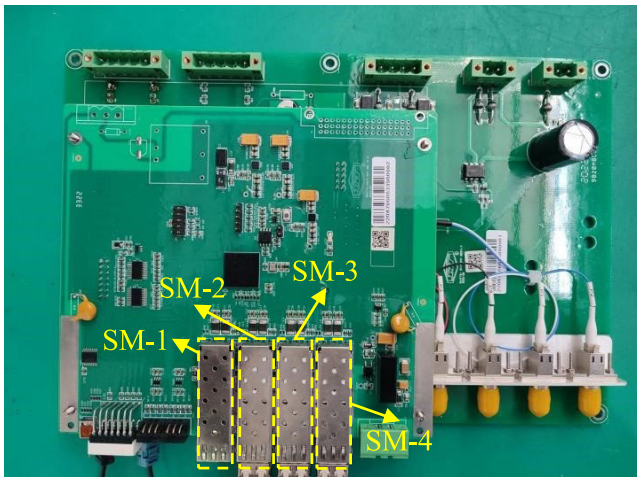

FIGURE 6. Fiber optic intensive SCE board receiving valve control instructions, where four optical modules are used for fault tolerance.

After obtaining a well-trained model, the performance of the fault diagnosis model is verified using the test set.

IV. EXPERIMENTAL SETUP AND RESULT DISCUSSION

A. EXPERIMENTAL SETUP

1) EXPERIMENTAL DEVICE

As a key component in flexible DC transmission systems, flexible converter valve equipment directly determines the operational performance of the entire flexible DC transmission system. The prototype of flexible converter valve equipment is shown in Figure 5. The core and bottom boards of the SCE board are shown in Figure 6. Each SCE board has four pairs of optical modules which are sequentially numbered. For example, the four optical modules on the #1 SCE board are labeled SM-1, SM-2, SM-3, and SM-4.

It should be noted that by default, the valve control instructions are first received by Receiver 1. If there is a communication failure with Receiver 1, Receiver 2 is used, and so on. The sub-modules perform frame synchronization based on Receiver 1; if there is a fault with Receiver 1, frame synchronization is performed using Receiver 2. In addition, all transmitting ports forward commands by 10 μs in advance (based on frame synchronization) and then generate transmission enable signals, transmission voltages, and status information based on their positions.

In real-world applications, a modular multilevel converter architecture consists of six bridge arms, each series connected by multiple flexible DC transmission converter sub-modules (over 100 of which are diagnosed in the experiment). After the series connection, sub-modules communicate through networking and are divided into normal working sub-modules and redundant sub-modules. In the case of abnormalities in normal working sub-modules (e.g., internal optical module, communication fiber, and photoelectric conversion circuit), the redundant sub-modules will be used. Modular multilevel converter will trip if the number of redundant sub-modules is insufficient. Considering the high cost of directly diagnosing modular multilevel converters and the disadvantage of mining the fault characteristics of each sub-module, this study focuses on the faults of sub-modules.

2) FAULT TYPES

The primary fault types diagnosed in this paper include component fault, fiber optic fault, power voltage drop fault, and abnormal flow injection fault. The fault voltage data collected using an oscilloscope are shown in Figure 7. The experiment was conducted in an industrial testing environment. The component fault was introduced by intentionally inducing a malfunction in the control system or the valve itself. The fiber optic fault was introduced by disconnecting a section of the fiber optic cable. The power voltage drop fault was introduced using a variable power supply to gradually reduce the voltage supplied to the system. The abnormal flow injection fault was introduced by modifying the flow control settings to inject a higher or lower flow rate than normal into the system. Note that each fault experiment only covered one type of fault to avoid the coupling effect of other faults on the fault signal. Examples of voltage waveform patterns under normal and different fault conditions are shown in Figure 8, where the horizontal and vertical axes represent sample points and voltage values. As shown in Figure 8 (b), component fault can stem from various issues, e.g., loose connections, inadequate welding, excessive valve vibration, and equipment overheating, leading to abnormal voltage fluctuations. According to Figure 8 (c), fiber optic faults could encompass issues such as loosening, detachment, and breakage. The fiber optic might endure physical or environmental harm, like excessive bending, breakage, or scratching, leading to irregular signal transmission. Disconnecting a section of the fiber optic cable caused fiber optic faults, leading to continuous voltage signal gains. As shown in Figure 8 (d), power voltage drop faults may arise from unstable power supply voltage, power line faults, or power overload, resulting in voltage drops. Loose

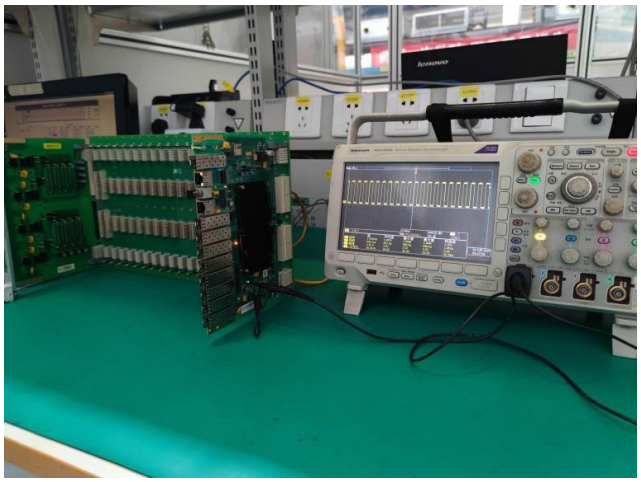

FIGURE 7. Acquiring voltage data of the experimental equipment under different fault conditions using an oscilloscope. The experimental environment is controlled, and fault experiments are conducted on different components with controlled variables to ensure that the fault data is only related to the set fault.

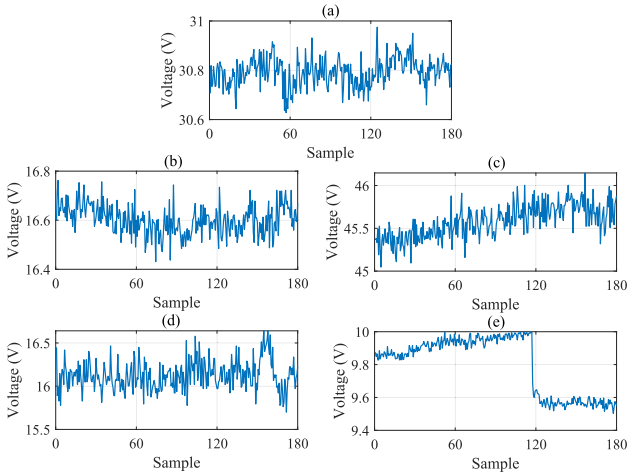

FIGURE 8. Voltage waveform under (a) No fault, (b) Component fault, (c) Fiber fault, (d) Voltage drop fault, and (e) Abnormal flow injection fault conditions.

or damaged connections could obstruct current transmission, also causing voltage drops. Additionally, defective power filters may play a role in voltage drops. As can be observed from Figure 8 (e), anomalous flow injection faults might originate from hardware malfunctions such as damaged network interfaces or processor failures. Misconfigurations or improper settings could redirect or inject anomalous flows, leading to voltage signal step changes. Figure 8 shows that the voltage data variation trends differ under different faults. However, the fluctuations in the voltage data are strong due to the influence of actual working conditions, and high-precision models are needed to deeply extract fault features.

3) DATA DESCRIPTION

Five scenarios are designed with no fault and each of the four types of faults. Among them, the data sample length in the no-fault scenario is 22,800, and those of the component fault, fiber optic fault, voltage drop fault, and abnormal flow injection fault scenarios are 102,300, 345,600, 526,100, and 125,300, respectively. As a newly developed equipment, the fault data generated during actual operation is insufficient. In addition, the cost of relevant fault diagnosis experiments is high due to high equipment costs. As a result, the fault data used for training is limited and unbalanced, requiring data augmentation.

Data augmentation methods for numerical data include data translation, data scaling, data noise, and overlapping sampling. The unique advantage of overlapping sampling is that it generates new samples by overlapping different samples, effectively increasing the data. Unlike traditional data augmentation methods, overlapping sampling does not require changing the shape and structure of the original data, allowing for increased data diversity without losing information. Therefore, it performs well in situations with limited data and sample imbalance. In this paper, each fault scenario is enhanced to 300 samples with 300 sequential data points based on the sample length of that scenario using different overlapping sampling steps, which also contributes to avoiding overfitting of the training model [30]. Maintaining the same sample size for each type of fault through data augmentation can avoid biased representation of fault categories. A total of 1,500 one-dimensional data samples with a size of 300×1 are obtained, with 80% of the samples used for training and 20% for testing. In the training samples, 20% are used to validate the model at this iteration.

4) EVALUATION METRICS

Using appropriate evaluation metrics to assess the effectiveness of the proposed fault diagnosis model is crucial. Therefore, the following well-known metrics are adopted [31]:

- Accuracy: As one of the most common evaluation metrics in classification tasks, accuracy represents the proportion of correctly predicted samples compared to the total number of samples, which is calculated as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (4)$$

- Precision: It indicates the proportion of true positive samples among all samples predicted as positive by the classifier, which is calculated as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (5)$$

- Recall: It represents the proportion of true positive samples predicted as positive by the classifier among all actual positive samples, which is calculated as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (6)$$

- F1-score: This metric combines precision with recall and is the harmonic mean of the two, which is calculated as

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (7)$$

Where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative results, respectively. These metrics complement each other, and considering them together can comprehensively evaluate the performance of a classifier.

5) OPERATING ENVIRONMENT AND PARAMETER CONFIGURATION

The experiments were implemented with Python 3.8 and Google's TensorFlow deep learning framework 2.3, which provide stable APIs and extensive community support and have a wide range of industrial applications and academic research support. The operating environment is Windows 11, and the simulation platform is equipped with a Core i7-1165G7 2.8 GHz CPU. The hyperparameters of the model are listed in Table 1. Among them, Adam optimizer is adopted, categorical cross-entropy is adopted as the loss function, and the mini-batch training method is adopted with a batch size of 16. Meanwhile, the number of iterations is set to 100, and the learning rate is set to 0.001. To avoid overfitting the training model, the Dropout function parameter is set to 0.3, and the early stopping strategy is applied. To estimate the variability and bias of the overlapping sampling operation, 10-fold cross validation is adopted to assess the stability and generalizability of the proposed method.

TABLE 1. Model hyperparameter settings.

Optimizer	Loss function	Iterations
Adam	classification cross-entropy	100
Training	Learning rate	Dropout
Min-batch batch_size=16	0.001	0.3

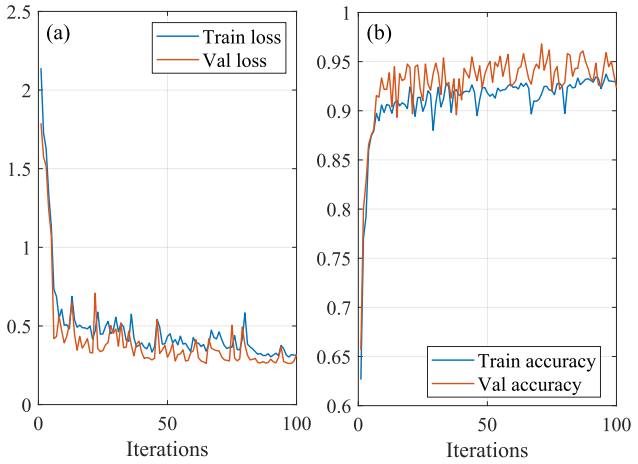

FIGURE 9. (a) Classification cross-entropy and (b) Diagnostic accuracy variations of the proposed method with the number of iterations.

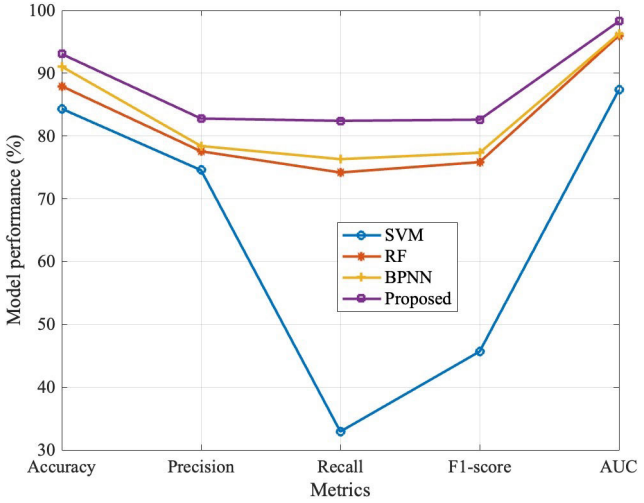

FIGURE 10. Performance of SVM, RF, BPNN, and the proposed method represented by blue, dark orange, orange yellow, and purple lines.

B. RESULT DISCUSSION

1) TRAINING PERFORMANCE

Based on the set model hyperparameters, the proposed fault diagnosis model is trained. Figure 9 shows the indicators of the loss function of the proposed method after 100 iterations, where the train loss and the val loss indicate the classification cross-entropy under the training and validating processes. As the number of iterations increases during model training, the loss value decreases and accuracy gradually improves. The gradual stabilization of both metrics indicates that the model gradually converges during training. In other words, the accuracy and classification cross-entropy of the model tend to stabilize during training, indicating that the fault diagnosis model has reached a relatively optimal state.

2) PERFORMANCE COMPARED WITH TYPICAL MACHINE LEARNING METHODS

Because the device is newly developed, the fault diagnosis method is compared with other methods in similar scenarios, including SVM in [15], RF in [13], and BPNN in [14]. Figure 10 presents the comparative results, revealing that the proposed method consistently outperforms these benchmark techniques. This superiority is particularly notable due to the non-linear separability inherent in the fault samples analyzed in this study. Non-linear separability implies that the boundaries between different fault classes cannot be effectively modeled by linear methods like SVM without complex transformations or kernels. As a result, SVM shows noticeably lower performance compared to RF and BPNN, which are more adept at capturing complex, non-linear relationships within the data. Furthermore, the results underscore the importance of selecting appropriate methods tailored to the specific characteristics of fault data. In this case, the proposed method demonstrates robustness and efficacy in handling the complexities of non-linearly separable fault patterns, thereby surpassing traditional SVM approaches.

3) PERFORMANCE OF DIFFERENT NETWORK STRUCTURES

Proposed method is also compared with several other methods with different network structures and attention modules under the same parameter configuration. Table 2 shows the model structures of the different methods. For

TABLE 2. Model structures of different methods, where "+" denotes serial connection. Drouput and Softmax operations are used in all models for Preventing model overfitting and calculating the confidence level for each fault type.

Model	Structure
Model 1 (M1)	3 layers of DSC+GAP+Dropout+Softmax
Model 2 (M2)	3 layers of BiLSTM+GAP+Dropout+Softmax
Model 3 (M3)	3 layers of BiGRU+GAP+Dropout+Softmax
Model 4 (M4)	1 layer of DSC+2 layers of LSTM +GAP+Dropout+Softmax
Model 5 (M5)	1 layer of DSC+2 layers of GRU +GAP+Dropout+Softmax
Model 6 (M6)	**1 layer of DSC+2 layers of BiLSTM +GAP+Dropout+Softmax**
Model 7 (M7)	1 layer of DSC+SE+2 layer of BiLSTM +GAP+Dropout+Softmax
Model 8 (M8)	1 layer of DSC+CBAM+2 layers of BiLSTM +GAP+Dropout+Softmax
Model 9 (M9)	1 layer of DSC+JAM+2 layers of BiLSTM +GAP+Dropout+Softmax
Proposed	**1 layer of DSC+TCJAM+2 layers of BiLSTM + GAP+Dropout+Softmax**

instance, "1 layer of DSC + SE + 2 layers of BiLSTM + GAP + Dropout + Softmax" in Model 7 indicates that its network structure is a DSC operation in the first layer, a Squeeze-and-Excitation (SE) attention operation in the second layer, two BiLSTM operations in the third and fourth layers, GAP and Drouput operations in the fifth layer, and a Softmax operation in the final layer.

Models 1 to 6 do not contain attention modules and are used to compare the fault diagnosis accuracy using only one-dimensional depth-wise convolutional layers (DSC), only BiLSTM layers, only BiGRU layers, and three hybrid models. Models 7 to 9 and the proposed method are all based on a model combining DSC and BiLSTM (M6). After the DSC layer, they each incorporate the SE attention module [32], the Convolutional Block Attention Module (CBAM) module [33], the Joint Attention Module (JAM) module [34], and the Time-Channel Joint Attention Module (TCJAM) module (the proposed method).

Figure 11 provides a detailed insight into the feature extraction process of fault-related data within the proposed fault diagnosis model. Initially, the time (Figure 11(a)) and frequency (Figure 11(b)) domains of the original signal illustrate the raw data characteristics before undergoing processing. Subsequently, in Figure 11(c) and (d), after passing through the DSC layer and the TCJAM layer, fault-related data features are highlighted within yellow rectangles. These features become notably more pronounced compared to the original frequency signal after DSC operation, effectively reducing noise artifacts generated during the convolution process. The combination of DSC and TCJAM layers enhances the distinction between fault-related and fault-free features, thus amplifying the signal-to-noise ratio crucial for accurate fault detection. This approach not only improves the diagnostic precision but also ensures robustness against the typical environmental and operational variability in flexible converter valve equipment.

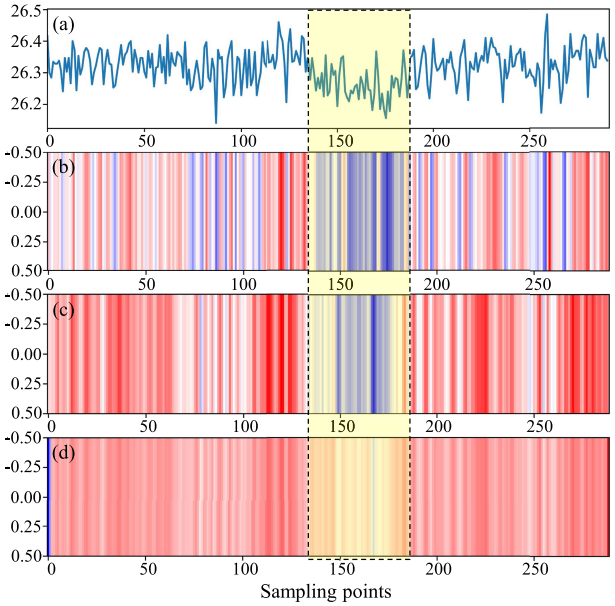

FIGURE 11. Significant polarization of data features after feature extraction from original signal (shown as (a) time domain and (b) frequency domain) to (c) DSC layer and (d) TCJAM layer.

The diagnostic accuracy of the above models is shown in Figure 12(a), where the box plots represent the diagnostic accuracy of 10 repeated experiments, and the lines represent the means of the diagnostic accuracy of 10 repeated experiments. The results indicate that the proposed method consistently achieves the highest diagnostic accuracy in repeated experiments (with a mean of 93.05%, 0.57% higher than the next-best method). These results demonstrate that the proposed method can accurately diagnose the five types of faults in flexible converter valve equipment. Furthermore, the proposed method exhibits the smallest deviation in multiple experimental results (1.47%), indicating its strong robustness.

The diagnostic precision, recall, and F1-score of the above models are shown in Figures 12(b), 12(c), and 12(d). Precision indicates the proportion of samples predicted as certain faults among all test samples of that fault. In contrast, recall indicates the proportion of actual samples with certain faults among all test samples accurately predicted with that fault. These evaluation metrics are commonly used to assess the performance of fault diagnosis models under sample imbalance. As shown in Figures 12(b), 12(c), and 12(d), the proposed method's diagnostic precision, recall, and F1-score are significantly higher than those of the comparison methods. These results indicate that the proposed method can accurately diagnose the five types of faults in the flexible converter valve equipment, which is critical for maintaining operational reliability and minimizing downtime.

In practical terms, the high recall rate observed in Figure 12(c) is particularly noteworthy. It signifies the model's ability to minimize false negatives, thereby reducing the likelihood of undetected faults that could potentially lead to equipment damage or operational failures. This aspect is

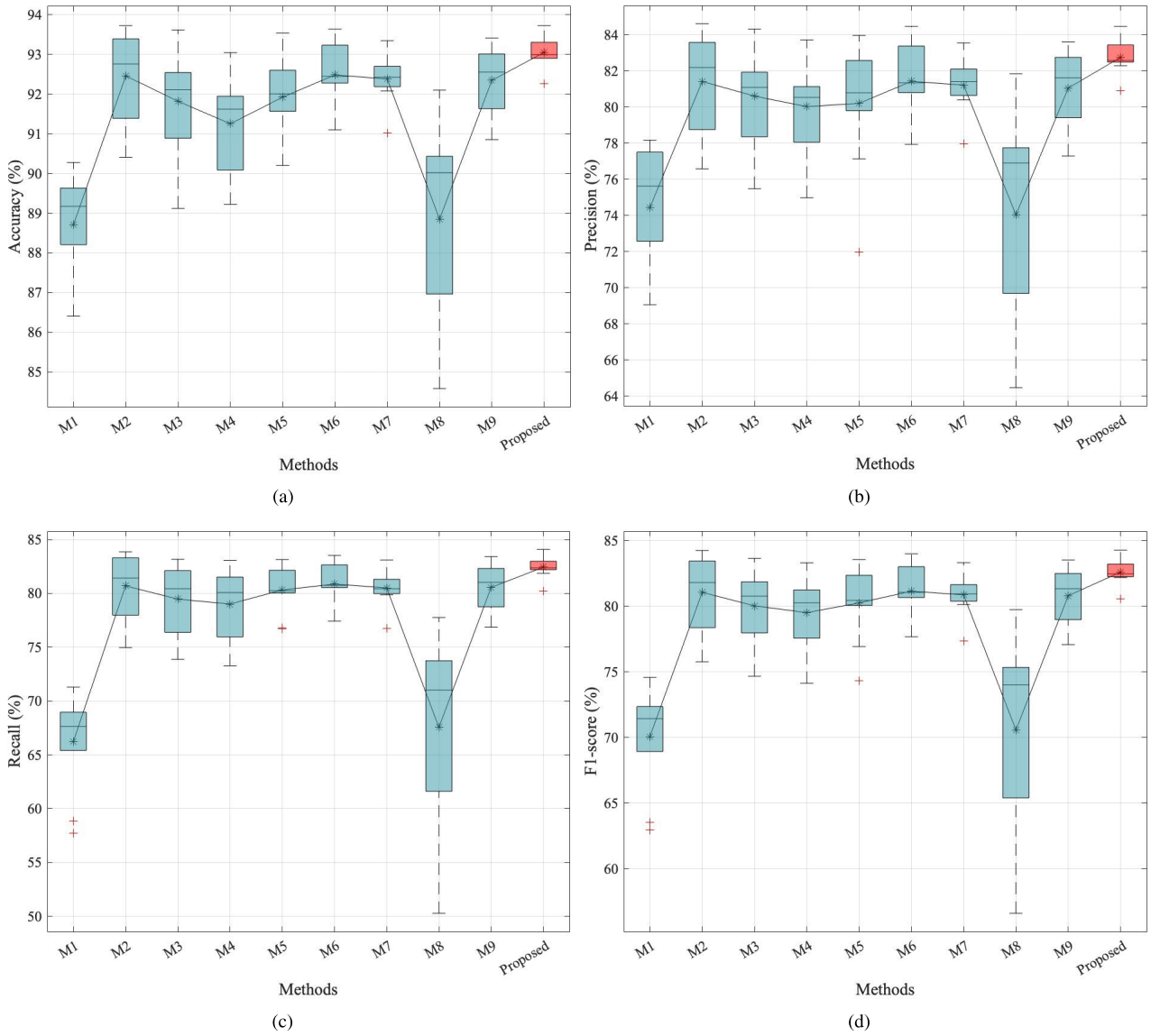

FIGURE 12. (a) Accuracy, (b) precision, (c) recall, and (d) F1-score of different methods, where the cyan box and red box represent compared methods and the proposed method, respectively. The box diagram is used to display the distribution of fault diagnosis results from multiple experimental groups. The degree of discreteness shown in the box diagram represents the robustness of the fault diagnosis method.

crucial in industries where system reliability and safety are paramount, such as in power grid stability and DC equipment operations. In addition, while false positives (Figure 12(b)) incur costs associated with unnecessary maintenance, they are less critical compared to the ramifications of missing true faults. Therefore, prioritizing a high recall rate aligns with operational priorities to ensure comprehensive fault detection without compromising system stability or incurring excessive operational costs.

In summary, the experimental results validate that the proposed fault diagnosis model excels in accuracy and reliability, making it well-suited for real-world applications in maintaining the operational integrity of flexible converter valve equipment. Its robust performance in both precision and recall positions it as a viable solution for industries requiring stringent fault detection capabilities.

4) PERFORMANCE UNDER DIFFERENT GAUSSIAN NOISE LEVELS

In the actual operation of power equipment, there are various noise sources, e.g., electromagnetic interference, environmental vibration, which may affect the quality of sensor data or signals. In addition, a good fault diagnosis method should be able to work effectively even under noise conditions that have not been seen during training. By simulating different Gaussian noise scenarios, it is possible to evaluate whether the proposed approach can be generalized to new noise conditions, rather than just performing well in clean environments during training. In order to evaluate the robustness, reliability, and generalization of the proposed method in noisy environments, this paper injects Gaussian noise of SNR 10 dB, SNR 20 dB, and SNR 30 dB in the test set, respectively. The experimental

TABLE 3. Performance of difference methods under noiseless, SNR 10 dB, SNR 20 dB, and SNR 30 dB gaussian noise levels.

Noise	Metric	M1	M2	M3	M4	M5	M6	M7	M8	M9	Proposed
Noise-less	Accuracy	88.71±2.31	92.46±2.05	91.82±2.70	91.25±2.03	91.92±1.72	92.48±1.38	92.38±1.37	88.85±4.27	92.35±1.49	**93.05**±0.79
	Precision	74.45±5.39	81.42±4.83	80.61±5.12	80.03±5.04	80.21±8.25	81.43±3.49	81.22±3.26	74.04±9.57	81.06±3.77	**82.76**±1.84
	Recall	66.22±8.49	80.70±5.73	79.47±5.60	79.01±5.73	80.32±3.60	80.87±3.45	80.51±3.76	67.55±17.27	80.56±3.69	**82.44**±2.24
	F1-score	70.08±7.13	81.06±5.29	80.03±5.36	79.51±5.39	80.26±5.94	81.15±3.47	80.87±3.51	70.58±13.97	80.81±3.73	**82.60**±2.04
SNR 10 dB	Accuracy	83.78±3.90	88.19±2.16	86.80±2.77	85.97±5.29	87.25±3.66	87.50±3.52	87.15±2.25	85.36±4.99	87.60±2.42	**88.57**±1.79
	Precision	69.92±5.30	78.23±1.78	77.21±5.98	77.22±11.99	77.58±7.42	76.87±5.75	75.72±4.82	69.40±15.42	78.02±4.41	**79.19**±4.41
	Recall	59.54±9.76	77.98±3.01	77.18±6.30	72.52±19.96	76.54±5.69	76.88±6.43	74.78±8.02	65.07±20.49	77.58±3.02	**79.06**±3.82
	F1-score	64.25±8.01	78.10±3.76	77.19±6.16	74.57±16.25	77.03±6.21	76.87±6.59	75.24±6.47	66.84±17.32	77.79±3.59	**79.12**±2.93
SNR 20 dB	Accuracy	85.55±3.51	90.09±1.82	88.16±2.28	86.05±3.26	89.36±2.47	89.71±2.00	89.59±1.61	86.69±3.07	90.01±0.95	**90.18**±1.77
	Precision	71.93±6.56	80.44±2.46	79.71±5.42	74.19±7.98	77.57±5.63	80.71±5.40	79.10±5.48	72.06±12.26	79.76±1.34	**80.80**±1.84
	Recall	61.57±7.69	80.20±1.95	79.06±6.85	71.68±10.22	77.36±5.00	80.00±4.84	78.77±6.34	66.04±11.96	79.32±1.99	**80.29**±1.97
	F1-score	66.22±6.22	80.32±2.20	79.38±6.14	72.84±9.19	77.46±5.31	80.35±5.12	78.93±5.91	68.78±12.32	79.54±1.67	**80.54**±1.90
SNR 30 dB	Accuracy	88.48±2.41	92.24±1.61	90.55±1.96	88.35±3.24	91.89±2.15	92.07±2.92	92.44±2.08	91.07±2.50	92.46±1.82	**92.85**±1.54
	Precision	74.15±6.24	81.54±3.44	81.81±5.34	76.29±7.87	79.66±3.02	80.84±7.49	80.99±5.59	77.64±6.14	80.96±3.71	**83.60**±2.31
	Recall	65.21±8.37	**81.39**±3.01	81.04±5.85	73.87±10.64	79.19±5.59	79.28±7.54	79.25±7.99	66.72±15.73	80.60±2.94	80.59±2.20
	F1-score	69.39±7.50	81.46±3.22	81.41±5.59	75.02±9.32	79.42±5.06	80.05±7.51	80.10±6.83	71.76±15.91	80.78±3.32	**82.00**±2.04

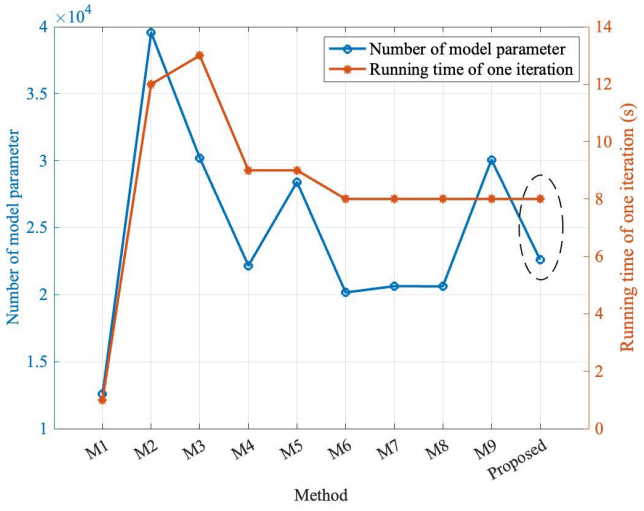

FIGURE 13. Parameter quantities and iteration time required for different methods, where the blue line and dark orange line represent the number of model parameter and running time of one iteration.

results are shown in Table 3. The results show that the proposed method can achieve the highest average accuracy in various Gaussian noise environments. In different noise environments, compared to the sub-optimal method, the proposed method has an average improvement of 0.42%, 1.18%, 0.53%, and 0.83% in Accuracy, Precision, Recall, and F1-score, respectively. The above results indicate that the proposed method effectively distinguishes fault modes and noise through attention modules, thereby enhancing the model's generalization ability in different noise environments. In addition, the deviation from multiple training processes is low, indicating that the diagnostic results of the proposed method have high consistency and stability.

In summary, high robustness and stability ensure the reliability and sustainability of the proposed method in practical industrial environments, even in the face of constantly changing noise conditions or data quality changes. In addition, stable multiple training results mean that the time cost of tuning and validation can be reduced, thereby accelerating the deployment and optimization process of fault diagnosis systems.

5) METHOD EFFECTIVENESS

Figure 13 shows the parameter sizes and the time required for one iteration of the ten fault diagnosis models. The blue line represents the model parameter size and the orange line represents the runtime for one iteration. The results show that the proposed method has a moderate model parameter size (22,616 parameters), which ensures sufficient model capacity without excessive computational overhead. Similarly, the proposed method has a moderate runtime for one iteration (approximately 8 seconds), which demonstrates efficient computational performance suitable for real-time or near-real-time fault diagnosis applications. In addition, it can be concluded that the proposed method has the highest diagnostic accuracy compared to other methods, and that the size of the model parameters and the training time are moderate. The combination of moderate model complexity, efficient computational performance, and superior diagnostic accuracy underscores the practical viability and effectiveness of the proposed method in real-world fault diagnosis scenarios. These attributes not only enhance the applicability of the method in diverse industrial contexts but also contribute to its potential scalability and usability in complex operational environments.

V. CONCLUSION

To address the limitations of traditional fault diagnosis methods for flexible converter valve equipment, this study presents a deep learning and attention mechanism-based fault

diagnosis method. Firstly, 1DCNN and BiLSTM are utilized to learn the temporal features in the data. Additionally, the time-channel joint attention module is proposed to achieve adaptive weight adjustment in the model, thus better capturing key information and reducing noise interference. To validate the effectiveness of the proposed method, various fault data are collected on flexible DC transmission system valves in experiments. To facilitate model training and testing, data augmentation is performed through overlapping sampling to balance the samples. The results indicate that the deep learning-attention mechanism-based method can accurately diagnose four types of faults in the flexible converter valve equipment of DC transmission systems. Compared with typical machine learning methods and different network structures, the proposed method demonstrates better performance with an average accuracy of 91.16%, an average precision of 81.59%, an average recall of 80.59% and an average F1-score of 81.07%. In different Gaussian noise levels, the proposed method almost has the best performance with an average improvement of 0.42%, 1.18%, 0.53%, and 0.83% in accuracy, precision, recall, and F1-score compared with the sub-optimal method, respectively. After multiple tests, the accuracy deviation rate and F1-score deviation rate of the proposed method do not exceed 1.79% and 2.93%, respectively, which is lower than the deviation values of most comparison methods. The above results indicate that the proposed method enable more accurate and stable diagnosis of various faults.

Although the proposed method has achieved good fault diagnosis results in the experimental environment, its performance may decrease due to real-world factors, e.g., system complexity (more than 22,000 model parameters and 8 seconds training time for one iteration) and uncertainty during long-term operation. Future research will focus on diagnosing faults after the series connection of multiple flexible converter sub-modules in DC transmission systems. In addition, improving the interpretability of the fault diagnosis model can help understand the model's decision-making process, build trust and reliability, and correct faults. To achieve this goal, model simplification and interpretable models are suggested.

REFERENCES

[1] G. M. Ud Din, N. Husain, A. Yahya, and Z. A. Arfeen, "Mathematical analysis of advanced high voltage direct current transmission systems," in *Proc. Int. Conf. Appl. Eng. Math. (ICAEM)*, Aug. 2021, pp. 61–66.

[2] T. Zhang, X. Yan, R. Zhang, Q. Ye, and J. Ma, "Distributed architecture of power grid asset management and future research directions," *IEEE Access*, vol. 10, pp. 57588–57595, 2022.

[3] C. Zhang, S. Zhao, Z. Yang, and Y. He, "A multi-fault diagnosis method for lithium-ion battery pack using curvilinear Manhattan distance evaluation and voltage difference analysis," *J. Energy Storage*, vol. 67, Sep. 2023, Art. no. 107575.

[4] C. Zhang, Y. He, T. Yang, B. Zhang, and J. Wu, "An analog circuit fault diagnosis approach based on improved wavelet transform and MKELM," *Circuits, Syst., Signal Process.*, vol. 41, no. 3, pp. 1255–1286, Mar. 2022.

[5] V. D. Phan, C. P. Vo, H. V. Dao, and K. K. Ahn, "Robust fault-tolerant control of an electro-hydraulic actuator with a novel nonlinear unknown input observer," *IEEE Access*, vol. 9, pp. 30750–30760, 2021.

[6] G. Li, W. Liu, T. Joseph, J. Liang, and Z. Song, "Double-thyristor-based protection for valve-side single-phase-to-ground faults in HB-MMC-based bipolar HVDC systems," *IEEE Trans. Ind. Electron.*, vol. 67, no. 7, pp. 5810–5815, Jul. 2020.

[7] W. Liu, G. Li, J. Liang, C. E. Ugalde-Loo, C. Li, and X. Guillaud, "Protection of single-phase fault at the transformer valve side of FB-MMC-based bipolar HVdc systems," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8416–8427, Oct. 2020.

[8] K. Xiao, Z. Wang, M. Peng, Y. Zou, X. Yan, and L. Xu, "Analysis and optimization strategy of dual slave fault in flexible-HVDC valve control system," in *Proc. 3rd Int. Conf. Energy Eng. Power Syst. (EEPS)*, Jul. 2023, pp. 578–581.

[9] A. A. Mamun, M. Sohel, N. Mohammad, M. S. H. Sunny, D. R. Dipta, and E. Hossain, "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," *IEEE Access*, vol. 8, pp. 134911–134939, 2020.

[10] R. Jiao, B. H. Nguyen, B. Xue, and M. Zhang, "A survey on evolutionary multiobjective feature selection in classification: Approaches, applications, and challenges," *IEEE Trans. Evol. Comput.*, early access, Jul. 5, 2023, doi: 10.1109/TEVC.2023.3292527.

[11] S. Ghashghaei and M. Akhbari, "Fault detection and classification of an HVDC transmission line using a heterogenous multi-machine learning algorithm," *IET Gener., Transmiss. Distrib.*, vol. 15, no. 16, pp. 2319–2332, Aug. 2021.

[12] Y. Ye, J. Zheng, and F. Mei, "Research on UPFC fault diagnosis based on KFCM and support vector machine," in *Proc. 14th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2019, pp. 650–655.

[13] P. Movahed, S. Taheri, and A. Razban, "A bi-level data-driven framework for fault-detection and diagnosis of HVAC systems," *Appl. Energy*, vol. 339, Jun. 2023, Art. no. 120948.

[14] L. Shuang and W. Sihua, "Fault location of flexible DC system based on BP neural network," in *Proc. 8th Int. Symp. Adv. Electr., Electron., Comput. Eng. (ISAEECE)*, May 2023, pp. 640–645.

[15] S. Zhou, J. Liao, and X.-j. Shi, "Kernel parameter selection of rbm-SVM and its application in fault diagnosis," *J. Electron. Meas. Instrum.*, vol. 9, pp. 69–74, 2014.

[16] C. Zhang, Y. He, L. Zuo, J. Wang, and W. He, "A novel approach to diagnosis of analog circuit incipient faults based on KECA and OAO LSSVM," *Metrology Meas. Syst.*, vol. 22, no. 2, pp. 251–262, Jun. 2015.

[17] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.

[18] C. Liu, F. Zhuo, and F. Wang, "Fault diagnosis of commutation failure using wavelet transform and wavelet neural network in HVDC transmission system," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.

[19] D. Wang, B. Wang, W. Zhang, C. Zhang, and J. Yu, "Fault location with high precision of flexible DC distribution system using wavelet transform and convolution neural network," *Frontiers Energy Res.*, vol. 9, Dec. 2021, Art. no. 804405.

[20] Y. Han, W. Qi, N. Ding, and Z. Geng, "Short-time wavelet entropy integrating improved LSTM for fault diagnosis of modular multilevel converter," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7504–7512, Aug. 2022.

[21] S. Zhou, L. Qin, Y. Yang, Z. Wei, J. Wang, J. Wang, J. Ruan, X. Tang, X. Wang, and K. Liu, "A novel ensemble fault diagnosis model for main circulation pumps of converter valves in VSC-HVDC transmission systems," *Sensors*, vol. 23, no. 11, p. 5082, May 2023.

[22] Á. Encalada-Dávila, L. Moyón, C. Tutivén, B. Puruncajas, and Y. Vidal, "Early fault detection in the main bearing of wind turbines based on gated recurrent unit (GRU) neural networks and SCADA data," *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 6, pp. 5583–5593, Dec. 2022.

[23] Y.-K. Liu, W. Zhou, A. Ayodeji, X.-Q. Zhou, M.-J. Peng, and N. Chao, "A multi-layer approach to DN 50 electric valve fault diagnosis using shallow-deep intelligent models," *Nucl. Eng. Technol.*, vol. 53, no. 1, pp. 148–163, Jan. 2021.

[24] L. Hou, L. Liu, and G. Mao, "Machine fault diagnosis method using lightweight 1-D separable convolution and WSNs with sensor computing," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–8, 2022.

[25] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023.

[26] C.-S. Wu, C.-H. Chen, C.-H. Su, Y.-L. Chien, H.-J. Dai, and H.-H. Chen, "Augmenting DSM-5 diagnostic criteria with self-attention-based BiLSTM models for psychiatric diagnosis," *Artif. Intell. Med.*, vol. 136, Feb. 2023, Art. no. 102488.
[27] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[28] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," *IEEE Access*, vol. 7, pp. 51522–51532, 2019.
[29] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis," *Future Gener. Comput. Syst.*, vol. 115, pp. 279–294, Feb. 2021.
[30] M. Molinier and J. Kilpi, "Avoiding overfitting when applying spectral–spatial deep learning methods on hyperspectral images with limited labels," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 5049–5052.
[31] S. U. Jan, Y. D. Lee, and I. S. Koo, "A distributed sensor-fault detection and diagnosis framework using machine learning," *Inf. Sci.*, vol. 547, pp. 777–796, Feb. 2021.
[32] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," 2018, *arXiv:1805.10180*.
[33] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.
[34] H. Wang, Z. Liu, D. Peng, and Y. Qin, "Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5735–5745, Sep. 2020.

JIANBAO GUO received the B.S. and M.S. degrees in power system and automation from Shandong University, China. Currently, he is the Senior Manager of the Electric Power Research Institute, EHV Maintenance & Test Center of China Southern Power Grid. He has authored or co-authored several journals/conference papers published in *The Journal of Engineering*, *Energies*, and IEEE 2023 Sixth ICET Proceedings. His research interests include conventional DC transmission engineering control and protection technology, flexible DC transmission engineering control and protection technology, and DC transmission engineering testing and debugging technology.

HANG LIU received the B.S. and M.S. degrees in power system and automation from North China Electric Power University, China. Currently, he is an Engineer with the Electric Power Research Institute, EHV Maintenance & Test Center of China Southern Power Grid. His research interests include conventional DC transmission engineering control and protection technology, flexible DC transmission engineering control and protection technology, and large-scale new energy DC transmission technology.

LEI FENG received the B.S. and M.S. degrees in power system and automation from the South China University of Technology, China. Currently, he is an Engineer with the Electric Power Research Institute, EHV Maintenance & Test Center of China Southern Power Grid. His research interests include conventional DC transmission engineering control and protection technology, flexible DC transmission engineering control and protection technology, and large-scale new energy DC transmission technology.

LIFENG ZU received the M.S. degree in power system and automation from the University of Shanghai for Science and Technology, China. Currently, he is the Software Development Manager of XJ Electric Flexible Transmission Company. His research interests include flexible DC transmission, high-voltage DC transmission, and new energy grid connections.

TAIHU MA received the M.S. degree in power system and automation from Southwest Jiaotong University, China. Currently, he is a Software Development Engineer with XJ Electric Flexible Transmission Company. His research interests include flexible DC transmission, high-voltage DC circuit breakers, and DC distribution networks.

XIAOLE MU received the B.S. degree in power system and automation from Wuhan University, China. Currently, he is a Software Development Engineer with XJ Electric Flexible Transmission Company. His research interests include flexible DC transmission and converter valve control software design.